A MODEL OF FRICATION NOISE SOURCE BASED ON DATA FROM FRICATIVE CONSONANTS IN VOWEL CONTEXT

P. Badin, K. Mawass and E. Castelli

Institut de la Communication Parlée, Grenoble, France

ABSTRACT

Aerodynamic and acoustic parameters have been measured for a subject uttering reiterant [pVFV] non-sense words at various loudness levels. Using *ensemble averaging* and *simplified inverse filtering* techniques, SPL and overall spectral tilt variations have been determined. A model of relations between these characteristics of the noise source and the pressure drop and minimum area at the constriction has been established by multilinear regression.

INTRODUCTION

The rapid growth of interest for articulatory synthesis calls for good articulatory *plants*, i.e. models capable of faithfully reproducing the articulatory, aerodynamic and acoustic behaviour of the human speech apparatus. In this domain, knowledge on noise excitation sources is still badly needed.

A recent study [1] has established relationships between noise source characteristics (SPL, overall spectral tilt) and aerodynamic characteristics (pressure drop and minimal area at the oral constriction), from data acquired for a subject sustaining voiceless fricatives. The aim of the present study was thus to extend this noise source model to fricatives in vowel context.

EXPERIMENTAL SET-UP AND CORPUS

Since the constriction area can be determined from the volume velocity and the pressure drop across the constriction, the set-up included a circumferentially vented wire-screen pneumotachograph, known as Rothenberg Mask to measure the volume velocity. In addition, the mask was equipped with a small polyethylene tube inserted through the lips at the mouth corner, and connected to a pressure transducer. The tube was running on one side of the mouth between the cheek and the gum, up to about 1 cm downstream the limit between the soft and the hard palates, in order to allow measurements of the pressure upstream a possible palatal-alveolar constriction. An electret microphone was placed at a distance of approximately 10 cm from the lips, and at an angle of 45° from the subject sagittal plane. The three signals delivered by the microphone, the pressure transducer, and the volume velocity transducer, were directly digitised at 12 kHz and stored on a computer disk.

The corpus consisted of reiterant [pVFV] non-sense words repeated about 10 times on one expiratory breath by one subject. The fricatives [f s ʃ] were combined with three symmetric vowel contexts [a i u], leading to 9 different items. Following the strategy already used in [1], each item was uttered at 18 different loudness levels, resulting in a SPL range of approximately 30 dB.

DATA PROCESSING

Segmentation

For each [pVFV] repetition, VTT (Voice Termination Time) and VOT (Voice Onset Time) instants were manually detected. The aim was to extract the largest possible portion of the fricative, including both transitions in and out of the fricative, but taking care to exclude any portion where voicing could be present, in order to ensure that the spectral characteristics of the signal would not be modified by the voice source. This resulted in nine fairly similar signals for each set of repetitions (the first repetition was systematically discarded). The length of the fricative segments thus defined varied between 190 and 300 ms over the whole corpus, with smaller variations for each context. For each of the nine repetitions, 20 measurement windows of 10 ms length were then uniformly distributed over the whole fricative segment, with some overlap if necessary, in order to allow the time alignment of the nine repetitions of each item.

Pressure drop and cross-sectional area at the constriction

The pressure drop Δp across the constriction is assumed to be very close to the Intra-Oral Pressure, and the volume velocity at the lips very close to that at the constriction. The Δp and volume velocity signals were first low-pass filtered at 80 Hz with a zero phase filter, and then used to determine the cross-sectional area A_c by means of the *orifice equation* [2]. Finally, Δp and A_c were averaged in each of the 20 measurement windows defined above. Fig. 1 gives an example of Δp and A_c trajectories obtained for [paʃa].
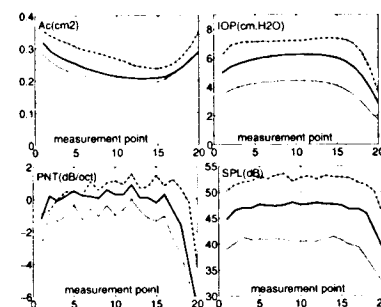

Fig. 1 – Example of time-aligned trajectories (solid line: medium; dotted line: soft ; dashed line: loud).
Top left: A_c; top right: Δp; bottom left: TLT; bottom right: SPL.

SPL and overall spectral tilt variations

In order to study the spectral variations of the fricatives in vowel context, two techniques have been combined: *ensemble averaging* [3] and *simplified inverse filtering* [1].

The spectrum of fricative sounds can be reliably determined by time averaging spectra computed from consecutive time windows throughout the duration of the fricative, if a segment of sufficient duration is available (about 100 ms). However, this technique can not be applied to the rapidly changing spectra in the transitions in and out of the fricative. Therefore, the alternative *ensemble averaging method* has been used (cf. [3]). In this case, averaging was done not *over* time, but at the *same* time across the ensemble of nine repetitions of the same fricative segment. Finally, for each item, 360 spectra (18 levels × 20 measurement

points) were obtained by ensemble averaging, cumulating a time duration of 20×10 ms. These spectra consist of 61 frequency bins of 100 Hz width.

As measuring directly vocal tract noise sources is impossible, indirect techniques have to be employed. Inverse filtering is widely used to determine voice source for vowel configurations, but such a technique is difficult to implement – and not very reliable – for fricatives [4]. Therefore, a procedure of *simplified inverse filtering* was used (cf. [1]). This procedure does not yield absolute spectral characteristics of the source spectrum, but provides an estimation of the variation of the overall spectral tilt of this spectrum as a function of speech effort. We have verified by simulation that small variations of the constriction area have little influence on the acoustic transfer function overall spectral tilt: the assumption that the variation of spectral tilt of the radiated sound is due to that of the source spectrum seems thus valid.

Simplified inverse filtering has been applied to each of the 9 items. First, third-octave spectra have been determined from the linear spectra mentioned above. The average of the corresponding 360 spectra has been computed for each item, and differential spectra have been estimated as the difference between each spectrum and the average. These spectra are deemed to represent the effects of the variations of the source spectrum, whereas the average spectrum represents the cumulated contribution of the average source spectrum, of the vocal tract transfer function and of the radiation characteristics at the lips. They have been found approximately flat between 200 and 5000 Hz. Overall spectral tilts (henceforth TLT) have finally been estimated as the slopes, expressed in dB/Oct., of the regression lines fitting these differential spectra. Thirteen third-octave bins were retained: from 250 Hz (224-280 Hz) to 4000 Hz (3550-4500 Hz).

On the other hand, overall SPL was computed as the RMS average of the sound pressure in each of the 20 measurement windows. The resulting trajectories of SPL and TLT are exemplified in Fig. 1.

Note that the validity of the simplified inverse filtering method has been verified

[1], including the effect of the mask upon the radiated sound.

RESULTS

Aerodynamic variables

An example of trajectories of Δp and A_c is displayed in Fig. 1. The minimum of each A_c trajectory has been determined for each item, each level and each context. Fig. 2 shows the span of this minimum for the different context. These data fit well with similar data measured on the same subject (PB) for the same corpus, but at a medium level only [5]. The variance of the present data is higher, likely due to the variation of loudness level.
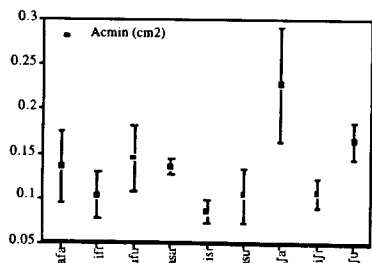
Fig. 2 – Minimum constriction area

In the previous study, a relatively strong correlation had been found between Δp and A_c. In the present corpus, these parameters are globally more independent, even though local correlations (at the level of one repetition) can be observed, as exemplified in Fig. 3 (V-shaped trajectories).
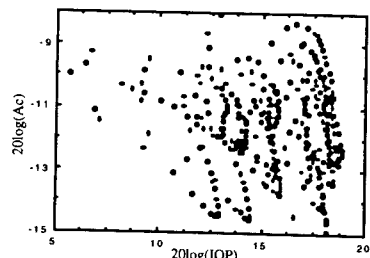
Fig. 3 – A_c vs. Δp for [pa∫a]

Acoustic variables

A correlation analysis has shown a high correlation between SPL and Δp, and a smaller, though not negligible, correlation between SPL and A_c. Similar results were found for the spectral tilts,

but with lower correlation coefficients. On the average, these correlations are lower than those found for the sustained fricatives in [1].

SOURCE VARIATION MODEL

Badin et al. [1] assumed that SPL and TLT can be expressed as:

$$SPL = k_1.\Delta p^p.A_c^q \qquad (1)$$
$$TLT = k_2.\Delta p^r.A_c^s \qquad (2)$$

where the exponents p, q, r and s, and the coefficients k_1 and k_2, can be determined by applying a multiple linear regression analysis to SPL and TLT as dependent variables, using $20\log_{10}(\Delta p)$ and $20\log_{10}(A_c)$ as independent variables. In order to extend the results on sustained fricatives to fricatives in vowel context, a similar analysis was performed for each of the nine contexts of the present corpus. Results are given in Fig. 4.
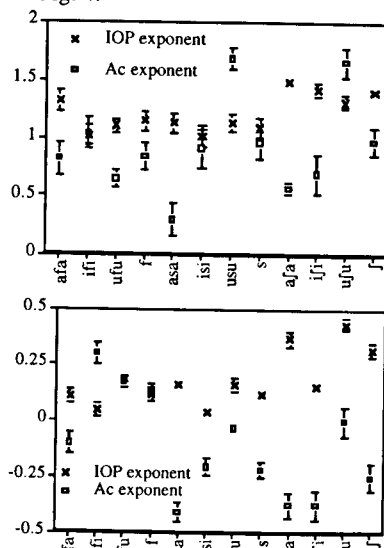
Fig. 4 – Δp and A_c exponents for SPL (top) and TLT (bottom), with the 95% confidence intervals

For each fricative, the exponents display a fairly large dispersion as a function of context. Computing exponents for each fricative, with the three vowel contexts pooled, did not lead to good results, because of unrealistic compensations between the influence of Δp and A_c. Instead, a single set of exponents has been derived for each

fricative, by averaging the exponents obtained for the three vowel contexts, and then determining the constant as the average of the residues, context by context, in order to take into account the level differences related to the three vowels. The resulting fit is less optimal, but presents the advantage of a unique model for each fricative. The coefficients obtained are shown in Table I.

	p	q	r	s
[f]	1.15	0.84	0.12	0.13
[s]	1.10	0.97	0.12	-0.22
[∫]	1.41	0.98	0.32	-0.25

Table I – Model exponents for the three fricatives

When comparing these exponents with those obtained for the sustained fricatives, the first striking observation is the relatively strong influence of A_c upon SPL (q close to 1). This may be ascribed to the fact that the span of A_c in the present data is larger, due to transitions in and out of the fricatives. It is also worth observing the negative correlation between TLT and A_c for [s ∫] whereas [f] exhibits a negative correlation (in fact for [ifi] and [ufu] only, cf. Fig. 3).

ASSESSMENT OF THE MODEL

It was first verified that the absolute error between the predicted and measured SPL values was about 2-3 dB. The error between TLT's is relatively larger, in proportion, about 0.5 dB/Oct. As for statistical models in general, we have observed that the prediction was less accurate for the data far from the centre of the distribution, i.e. the low values of Δp and the high values of A_c (up to 6 dB for SPL and up to 2 dB/Oct. for TLT).

Third-octave spectra have been finally re-synthesised as a function of Δp and A_c. For each single measurement point, SPL and TLT were first estimated, and then synthetic third-octave spectra were computed as the sum of the average spectrum established for corresponding context, and of the SPL and TLT variations. Fig. 5 displays the measured and synthetic spectra at 3 measurement points for [pa∫a]. The analysis of the results has also shown that the mean absolute error (i.e. the average of the absolute differences of the third-octave spectra in dB) between measured and synthetic spectra was about 2-3 dB, with occasional peaks up to 6-7 dB.
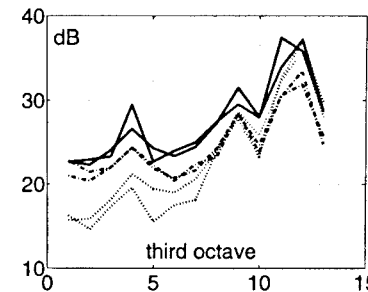
Fig. 5 – Example of 3 measured and re-synthesised spectra for [pa∫a].

CONCLUSIONS

We have presented a model of variation of noise source spectrum based on a corpus of fricatives in vowel context. The model predicts the variations of SPL and overall spectral tilt of the source spectrum as a function of the aerodynamic state in the vocal tract. This model will next be implemented in an articulatory synthesiser and thus tested in a complete articulatory synthesis scheme.

ACKNOWLEDGEMENTS

REFERENCES

[1] Badin, P., Shadle, C.H., Pham Thi Ngoc, Y., Carter, J.N., Chiu, W., Scully, C., & Stromberg, K. (1994). Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis. *ICSLP*, Yokohama, Japan, Vol.1, 163-166.
[2] Scully, C. (1986). Speech production simulated with a functional model of the larynx and the vocal tract. *Journal of Phonetics*, 14, 407-414.
[3] Shadle, C.H., Dobelke, C.U., and Scully, C. (1992). Spectral analysis of fricatives in vowel context. *J. de Physique IV*, Coll. C1, supp. au J. de Phys. III, vol.2, April 1992. Pages C1-295 to C1-298.
[4] Badin, P. (1991). Fricative consonants: acoustic and X–ray measurements, *Journal of Phonetics* 19, 397-408.
[5] Stromberg, K., Scully C., Badin, P., & Shadle, C.H. (1994). Aerodynamic patterns as indicators of articulation and acoustic sources for fricatives produced by different speakers. *Proc. of the Institute of Acoustics*, Vol. 16, Pt 5, pp 325-333.